

Introducing the ICPSR Metadata Documentation Portal

jay winkler (jmwink@umich.edu)

Mike Shallcross (shallcro@umich.edu)

February 16, 2024

International Love Data Week

What are Metadata?

*Information used to describe a work to enable
discovery and reuse.*

ALA Glossary of Library and Information Science

How does ICPSR use metadata?

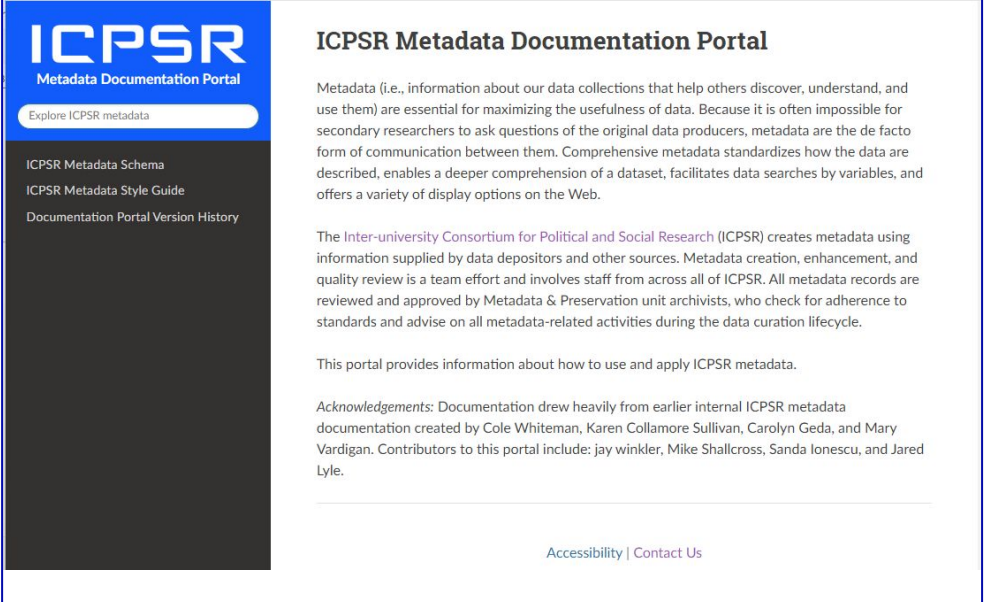
- Powers ICPSR search
 - Free Text
 - Faceting
- Machine-actionable exports
 - Datacite
 - Object Export API (forthcoming)

How does ICPSR generate metadata?

- “Silent collaboration”- Depositors provide information, curators sort through that information for valid metadata
 - Depositor-provided metadata
 - Information from documentation
 - Primarily compiled by curation staff

What is the ICPSR Metadata Documentation Portal?

- Website with relevant information about ICPSR metadata practices.
- Created for ICPSR staff and the public.
- <https://icpsr.github.io/metadata/>



The screenshot shows the ICPSR Metadata Documentation Portal. The header features the ICPSR logo and the text "Metadata Documentation Portal". Below the header is a navigation bar with a button labeled "Explore ICPSR metadata". The main content area is dark grey and contains links for "ICPSR Metadata Schema", "ICPSR Metadata Style Guide", and "Documentation Portal Version History". The right side of the page has a white background with the title "ICPSR Metadata Documentation Portal" and a paragraph explaining the importance of metadata. Below this is another paragraph about the ICPSR's metadata creation process, followed by a paragraph about the portal's purpose. At the bottom, there is an "Acknowledgements" section and a footer with links for "Accessibility" and "Contact Us".

ICPSR Metadata Documentation Portal

Metadata (i.e., information about our data collections that help others discover, understand, and use them) are essential for maximizing the usefulness of data. Because it is often impossible for secondary researchers to ask questions of the original data producers, metadata are the de facto form of communication between them. Comprehensive metadata standardizes how the data are described, enables a deeper comprehension of a dataset, facilitates data searches by variables, and offers a variety of display options on the Web.

The Inter-university Consortium for Political and Social Research (ICPSR) creates metadata using information supplied by data depositors and other sources. Metadata creation, enhancement, and quality review is a team effort and involves staff from across all of ICPSR. All metadata records are reviewed and approved by Metadata & Preservation unit archivists, who check for adherence to standards and advise on all metadata-related activities during the data curation lifecycle.

This portal provides information about how to use and apply ICPSR metadata.

Acknowledgements: Documentation drew heavily from earlier internal ICPSR metadata documentation created by Cole Whiteman, Karen Collamore Sullivan, Carolyn Geda, and Mary Vardigan. Contributors to this portal include: jay winkler, Mike Shallcross, Sanda Ionescu, and Jared Lyle.

[Accessibility](#) | [Contact Us](#)

Under the Hood...

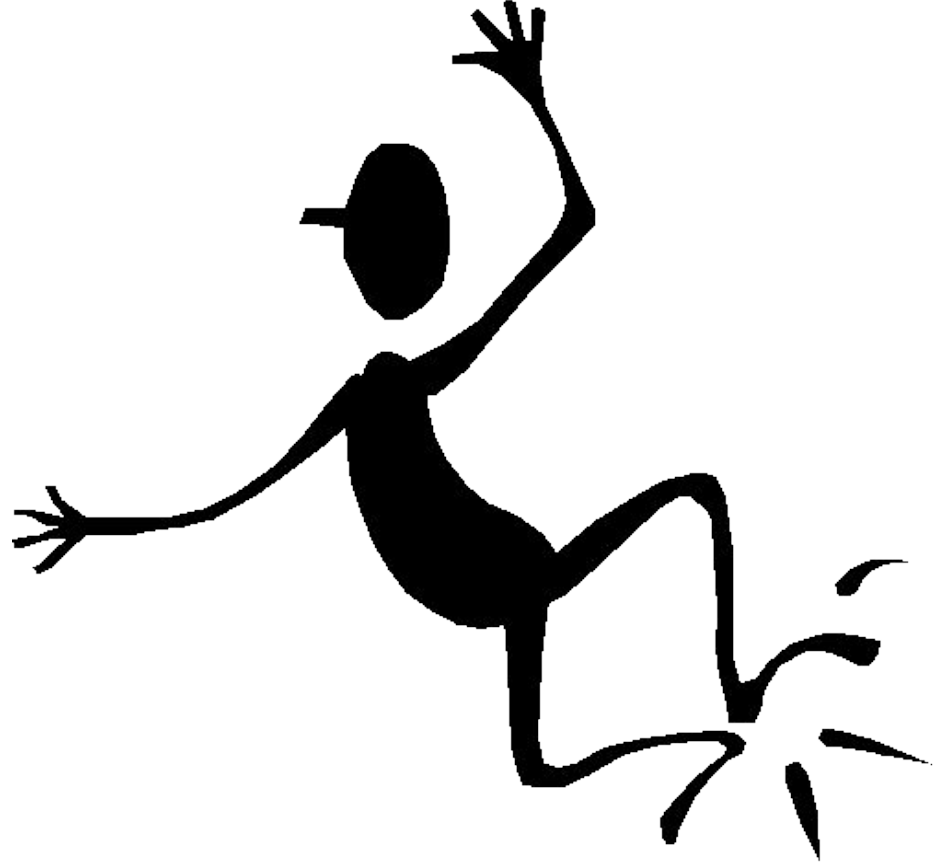
- Metadata documentation is maintained in a [JSON Schema](#) document in a [GitHub repository](#).
 - JSON Schema is an international standard
 - Used to write standardized rules for how information can be input and validated
- Automated workflows (“[GitHub Actions](#)”) convert documentation to HTML that is published using [GitHub Pages](#)

Why Create a New Metadata Portal?

- It was time: regular documentation reviews and updates are essential.
- Old documentation was spread across multiple locations; challenging to (a) find information and (b) keep internal and external versions in sync.
- Increase transparency and understanding: more detailed definitions, usage notes, and examples.
- Built on GitHub; easier to publish changes and track version history.

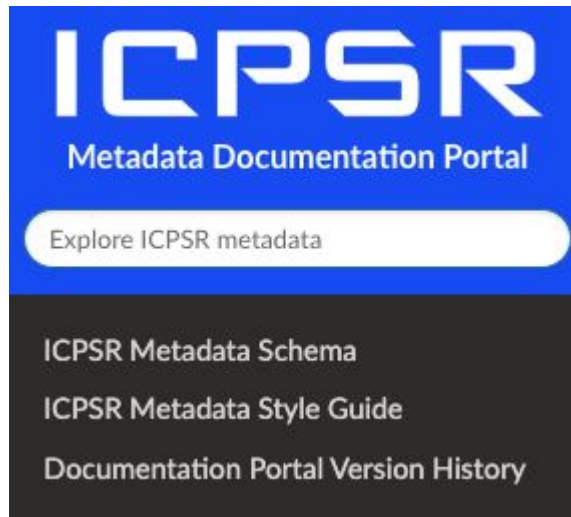
Some Things Stay the Same...

- Portal reflects ICPSR's 60+ years of experience.
- No significant changes to metadata definitions
- ICPSR continues to align with community standards, like DDI.



Portal Content: an Overview

- ICPSR Metadata Schema: information about all the fields on ICPSR study home pages.
- ICPSR Metadata Style Guide: grammar and usage conventions common to all metadata fields.
- Version History: record of changes to the documentation portal.



ICPSR Metadata Schema

The screenshot shows the ICPSR Metadata Documentation Portal. At the top is the ICPSR logo and the text 'Metadata Documentation Portal'. Below that is a search bar with the text 'Explore ICPSR metadata'. The main navigation menu is on the left, with 'ICPSR Metadata Schema' expanded. Under this menu, 'Metadata Elements: Detailed Information' is selected, and a red arrow points from it to the '1. Version' item in the table of contents.

- ICPSR Metadata Schema
 - Metadata Elements: Overview
 - Key for Metadata Element Entries
 - Metadata Elements: Detailed Information
 - 1. Version
 - 2. Version Date
 - 3. Original Release Date
 - 4. Title
 - 5. Alternate Title
 - 6. Link Title
 - 7. Link URL
 - 8. Principal Investigator
 - 9. Citation
 - 10. Distributor
 - 11. Digital Object Identifier (DOI)
 - 12. Funding Source

ICPSR Metadata Schema

Last updated: January 29, 2024 **1**

This is the metadata schema used to describe data collections at the Inter-university Consortium for Political and Social Research (ICPSR). These rules and definitions represent ICPSR's metadata practices and are intended to (a) assist ICPSR staff with metadata entry, and (b) help ICPSR users -- including data depositors and researchers accessing data -- understand how to use and interpret our metadata.

For a machine-actionable copy of this information, please see the [JSON Schema version](#). **2**

Metadata Elements: Overview

A red arrow points from the '1. Version' item in the navigation menu to the first row of the table.

Property	Required?	Repeatable?	Accepted Values	Description
Version	Yes	No	Number	The current version number for the data collection.
Version Date	Yes	No	Text	The date on which the current version of the data collection was released by ICPSR.
Original Release Date	No	No	Text	The date on which the data collection was originally released by ICPSR.
				The authoritative title of the data collection.

Metadata Element Definition Key (I)

- Description: Defines the metadata element and the information it is intended to convey.
- Required: Indicates whether the metadata element is mandatory ("Yes") or optional ("No"). Required elements must include at least one value.
- Repeatable: Indicates whether the metadata element may be repeated ("Yes") or if it may only occur once ("No").
- Accepted values: The type of values that may be used with the metadata element (i.e., text or numbers).

Metadata Element Definition Key (II)

- Controlled Vocabulary: Indicates the specific controlled vocabulary used to provide values for the metadata element ('N/A' indicates no controlled vocabulary is required).
- Usage Notes: Additional information about the nature, scope, and conventions for values that may be added to the metadata element.
- ICPSR Input Guidance: Information for ICPSR staff related to using internal tools and resources.
- Examples: Illustrates the element's usage with valid values.

Example Metadata Schema Entry

15. Subject Term

Description: A controlled list of terms used for indexing purposes to indicate what topics a data collection reflects. Subject Terms serve to increase the collection's discoverability by topic.

Required: Yes

Repeatable: Yes

Accepted Values: Text

Controlled Vocabulary: The [ICPSR Subject Thesaurus](#) and the [ICPSR Personal Names Authority List](#) are preferred sources.

Usage Notes: The [Library of Congress Subject Terms](#) is referenced when adding new terms to the ICPSR Subject Thesaurus.

ICPSR Input Guidance: Non-thesaurus terms can be submitted in the metadata editor and will be reviewed by the metadata librarian. If an ICPSR staff member submits a non-thesaurus term, the metadata librarian will gauge the necessity of this term, check it against the Library of Congress Subject Headings to see if a different related term should be used, and consider it for addition to the ICPSR thesaurus.

Examples:

```
[  
  "child care"  
]
```

```
[  
  "Social Security",  
  "crime",  
  "victimization"  
]
```

Subfields & Complex Metadata Elements

- Some metadata (like “Principal Investigator”) are actually comprised of multiple subfields.
- Subfields are related and grouped together (e.g., a PI will have a specific affiliation and ranked order)
- Subfield definitions use the same conventions.

8. Principal Investigator

Description: The entity or entities responsible for the data collection, presented in order of importance. A Principal Investigator (PI) may be a person or an organization. Each data collection requires at least one PI.

Required: Yes

Repeatable: Yes

Accepted Values: Multi-part element; see subfield definitions for more information.

Controlled Vocabulary: N/A

8.1. Subfields:

Property	Required?	Repeatable?	Accepted Values	Description
Name	Yes	No	Text	The name of the entity primarily responsible for the data collection.
Affiliation	No	No	Text	The organization with which an individual PI was affiliated at the time of a data collection's deposit at ICPSR.
Order	Yes	No	Number	The order of importance for the PIs associated with the data collection -- typically provided to ICPSR by the lead PI.

A Step Forward in Collaboration

What the Metadata Documentation Portal IS:

- A way for us to help each other make sure all studies are as findable and accessible as possible
- A way for you to understand why we might make changes to what was submitted

What the Metadata Documentation Portal IS NOT:

- Homework for homework's sake
- A major change to the rules surrounding metadata submission at deposit
- A way for us to “reject” metadata

Strengthening the “Silent Collaboration”

- Working from same starting point
- Better understanding of potential changes made to study metadata
- Venues for feedback to metadata team
- Ongoing transparency about changes to schema

Q/A

